Selecting ASPLOS Best Papers: A Guide

DAN TSAFRIR, Technion – Israel Institute of Technology, Israel
MADANLAL MUSUVATHI, Microsoft Research, United States

ASPLOS 2024 received a record of 922 submissions (spread across three review cycles), from which 194 papers were accepted. This record represents an increase of about 50% over 2023, which itself saw a similar increase over 2022. As a result, selecting the best papers proved especially challenging. In this document, we describe the methodology we employed—a process we believe is fair, upholds high standards of quality, and emphasizes consensus and expert judgment. We hope it will be useful for others.

CCS Concepts: • **General and reference** → **Evaluation**.

Additional Key Words and Phrases: Reviewing, computer systems conferences

ACM Reference Format:
Dan Tsafrir and Madanlal Musuvathi. 2026. Selecting ASPLOS Best Papers: A Guide. *ACM Trans. Comput. Syst.* 44, 3, Article 21 (August 2026), 7 pages. https://doi.org/10.1145/3815200

1 Introduction

Selecting the best papers for ASPLOS poses a challenge due to a dramatic increase in the number of submissions and a multi-cycle review structure. This article outlines the approach we employed in ASPLOS'24. We begin by summarizing the historical context and naming of the award (which has been inconsistent), motivating our decision to use a unified process across per-year cycles (Section 2). We then discuss the target number of awards and how we set flexible, yet principled bounds (Section 3). We next describe the criteria we used to identify and filter candidate papers based on preexisting reviews and consensus (Section 4).

We detail the formation of the award committee, emphasizing a data-driven, inclusive process for identifying the top reviewers in our program committee (consisting of 224 members), whom we invited to serve on the best paper selection committee (Section 5). In recognition of their contribution, we introduced a new "distinguished reviewer" award [1] in ASPLOS'24, (1) to acknowledge their excellent work up to that point, and (2) to reward their willingness to take on additional reviewing load beyond their regular duties. (The award was based solely on merit, regardless of whether recipients agreed to partake in the selection process.)

We then explain why and how we bootstrapped a separate HotCRP instance to manage reviewing (Section 6), describe the custom review form and evaluation criteria (Section 7), and outline the reviewer assignment process used to ensure fair and expert evaluations (Section 8). Finally, we conclude with a summary of the award selection meeting (Section 9) and our reflections on the overall process. We hope this documentation will serve as a useful reference for future program chairs.

Authors' Contact Information: Dan Tsafrir (corresponding author), Faculty of Computer Science, Technion – Israel Institute of Technology, Haifa, Haifa District, Israel; e-mail: dan@cs.technion.ac.il; Madanlal Musuvathi, Microsoft Research, Redmond, Washington, United States; e-mail: madanm@microsoft.com.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 0734-2071/2026/08-ART21
https://doi.org/10.1145/3815200

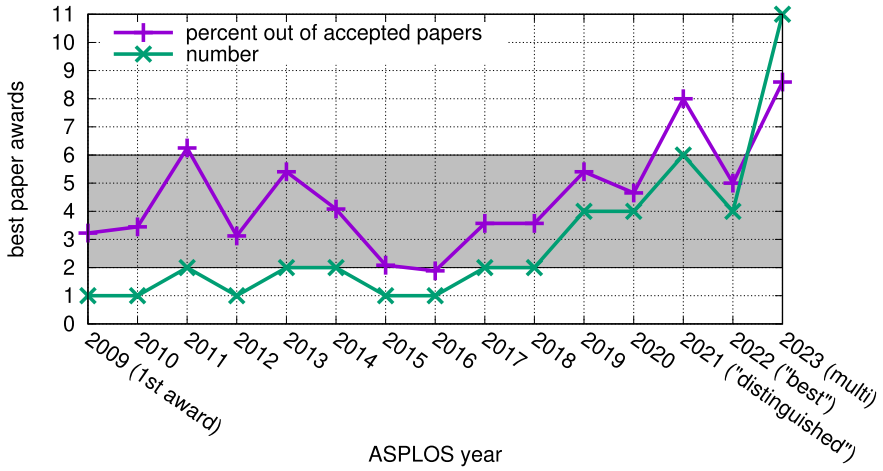


Fig. 1. The number of best paper awards handed in ASPLOS and their percentage out of the number of accepted papers. When considering percentages, two ASPLOS instances stand out as being outside the 2%–6% range, which is highlighted with a background color.

2 Historical Data and Naming of the Award

ASPLOS has been handing out best paper awards since 2009 [7]. The typical number of awards ranges between 1 and 4, constituting 2%–6% of the accepted papers. Figure 1 shows all the historical data of ASPLOS best paper awards, highlighting two exceptions. In 2021, the program chairs handed 6 awards to 8% of the accepted papers, naming them "distinguished papers," a term commonly used by SIGPLAN conferences with an implied meaning that up to 10% of the accepted papers can be awarded [2]. A year later, the program chairs returned to the traditional "best paper" name and the traditional statistics. Subsequently, in 2023, ASPLOS became a multi-cycle conference, and its program chairs opted to again use the "distinguished" term, awarding 11 papers in total through a per-cycle selection procedure that associated awardees with their cycle. Thus, there were two spring awards, five summer awards, and four fall awards in ASPLOS'23.

We are of the opinion that ASPLOS should select a name for the award and stick with it rather than alternate on a yearly basis. Frequent changes might be distracting, and we are worried that they might affect the perception of the award. Given that the award has historically been referred to as "best paper" and has been presented as such in all but two instances, it is arguably preferable to maintain this tradition. We do not see strong motivation that justifies the newer branding. Importantly, we do not see a connection between the name of the award and the number of times that it is handed. ASPLOS program chairs are free to award up to 10% of the accepted papers, regardless of the name.

In this context, given that ASPLOS is now multi-cycle, we would like to additionally claim that associating awards with specific ASPLOS cycles might detract from their prestige. This is not just a matter of naming; it is also a matter of substance: Being selected as the "best" from among all papers submitted throughout the year is more significant and meaningful than being the best within a single cycle. The achievement is greater.

Moreover, we would like to avoid situations in which an award paper in spring, for example, would not have been selected in summer due to much stronger competitors. In our opinion, the strongest papers should be selected regardless of their cycle. The specific timing of the submission should not be a factor, we believe. Consequently, in 2024, we decided to assess all candidate papers collectively through a unified paper selection process.

3 Target Number of Awards

Opinions vary with respect to the number of awards that should be handed per year. Some believe that a higher target of 10% of the accepted papers is preferable, as it is individually beneficial to more authors and thus to the community as a whole. Others contend that 10%—which amounts to nearly 20 in our case—is excessive, diluting the perceived value of the award and potentially exceeding the reading capacity of attendees who may treat the award as a signal for which papers to sample in order to get the "pulse" of the conference. Another drawback of targeting 10% is the increased reviewing load that must be shouldered by the selection committee.

We did not set a specific target award number. But we agreed that it should be at least 6, namely, 3.5% of the accepted papers, which is in the lower half of the highlighted area in Figure 1. We also agreed that physically handing awards at the conference to more than double this many could be awkward. In the review form of candidate papers, when providing context to the selection committee, we explained: (i) that single-cycle ASPLOS typically handed 2–4 awards per year (in Figure 1, this statement applies since 2017 if disregarding the two exception instances); and (ii) that ASPLOS'24 is approximately 3x bigger than older single-cycle ASPLOSes, so it can hand 3x more awards (which translates to a range of 6–12 or 3.5%–7%). We further noted that (iii) if the selection committee ends up feeling that more candidates are award-worthy, we may consider handing more awards.

4 Award Candidate Selection

Motivated by the procedure to select award papers that was suggested by Emery Berger [2] (employed in his capacity as PLDI'16 program chair), we only considered candidates without detractors. We further limited the candidate pool by determining that the number of "accept" or "strong accept" merit scores should be equal to or greater than the number of "weak accept" scores. In practice, this restriction implied no more than two "weak accept" scores. Applying this criterion resulted in 18 candidates out of the 170 accepted papers at the time the selection process was conducted. (Subsequently, after ASPLOS'24 took place, we additionally accepted 24 fall major revisions; see below.) There were representatives from all cycles albeit not uniformly distributed with respect to the number of accepted papers in each cycle.

The "best paper candidate" scores that reviewers associated with submissions throughout the 3 reviewing cycles via the review forms turned out to be a weak signal that did not eventually help us to identify our candidates. In retrospect, we do not think that this field is necessary. We contend that a positive consensus among the reviewers constitutes a stronger, more appropriate criterion.

We note in passing that our selection procedure disadvantages papers that underwent revisions, as they are unlikely to have no detractors and a plurality of strong supporters; otherwise, they would likely have been accepted without a revision. Notably, in ASPLOS'24, only a subset of the original reviewers re-reviewed revisions, so the scores of the other reviewers remain unchanged. Indeed, all of our 18 candidates did not undergo revisions, and the fall revisions that were under review while the selection process took place would not have been able to meet our above selection criterion.

In principle, we could have addressed this problem at the cost of increasing the reviewers' workload, by asking all of them to re-review. But we believe that doing so for this particular reason (a small chance to get an award) would have been unfair to reviewers. Moreover, arguably, it would have been unfair to authors of regular accepts, who did not get the opportunity to work for an additional six weeks to address reviewer feedback. Additionally, disregarding major revision improvements seems no different than disregarding the camera-ready improvements of our fall candidates due to the tight schedule. We are comfortable with this compromise and believe it is reasonable.

5 Assembling the Award Committee and Distinguished Reviewers

We assembled a best paper selection committee from the members of the ASPLOS'24 program committee (PC). Our committee consisted of three groups of members: (1) our eight vice program chairs (VCs), who, by the nature of their service, have a broad perspective of the ASPLOS'24 papers; (2) the three program/general chairs of ASPLOS'25 who served in our PC and, as such, are established members of the ASPLOS community; and (3) a group of 12 distinguished reviewers selected as described next. Because 2 of our VCs are also ASPLOS'25 chairs, our best paper selection committee consisted of 21 (8+3+12-2) members.

We decided that selecting the best ASPLOS papers is not just a service but also a privilege that should go to *distinguished* members of our PC who have excelled in their service during their year-long term. (Distinguished reviewers are formally recognized in the ASPLOS'24 website.) To identify the distinguished reviewers, we used a data-driven approach, hoping to avoid implicit bias in our selection. We used the following combination of signals computed across all three cycles: (1) the number of times a PC member received a "good review" rating from their fellow reviewers; (2) their contributions to reviews and discussions, approximated via word count; (3) the timeliness of their reviews and comments; and (4) our misbehavior notes (very few), which reduced ranking. Our metric identified 18 top-ranking reviewers [1], of which 12 graciously accepted our invitation to serve. To bound the workload and make it manageable, we determined that each committee member would not review more than five award candidates. There was no overlap between the authors of candidate papers and the award selection committee.

The aforementioned selection procedure [2] recommends appointing only senior researchers as members of the award selection committee. But after some consideration, we decided to also include our top-ranking junior colleagues, based on merit, and for diversity of opinion. A welcome side effect of this decision was that it increased the visibility of the junior members (notably through interacting with their more senior colleagues during the meeting), and it also provided them with some valuable experience, we believe.

Our list included the usual suspects, namely, established members whom we knew to be excellent reviewers. But we were also pleasantly surprised to meet individuals that were relatively junior or unfamiliar to us, doing excellent work in supporting ASPLOS through their dedicated service. While our committee was diverse in terms of seniority, expertise (the 3 ASPLOS disciplines), and geography, it was not gender diverse. Of the 18 reviewers we invited, 3 were women (matching the gender diversity of the PC), but all of them declined our invitation due to prior commitments, unfortunately.

6 Bootstrapping

We used a new HotCRP instance for best paper reviewing, which was necessitated for a number of reasons (such as having candidates from multiple cycles). HotCRP is not natively designed to allow regular users to bulk-upload submissions. Yet, manually uploading each candidate paper along with all of its metadata would have been time-consuming and error-prone. Thankfully, the HotCRP maintainer graciously helped us to bulk-upload the candidate papers with their metadata (while ensuring that emails were not sent to the authors). The metadata included topics and conflict of interest declarations. We nevertheless re-executed the conflict-finding scripts in case some more CoIs were created since the papers were submitted—nearly a year ago for some of them.

When bulk-uploading, we used the latest PDF versions available for each candidate. Camera-ready versions for fall papers were unavailable until the end of the best paper review process, which started in early March 2024 shortly after the fall author notification date, and ended in early April. Due to this timing constraint, we used the originally submitted PDF versions for these papers. In contrast, we used camera-ready versions for candidate papers from earlier cycles. We

concede that this policy could potentially bias the selection towards papers submitted in earlier cycles, although this did not seem to be the case in the eventual selection outcome. As of 2026, the number of ASPLOS yearly cycles is reduced from 3 to 2, so the schedule allows program chairs to use camera-ready versions only.

Unlike the per-cycle HotCRP instances, we configured the best paper instance to be more permissive: the papers were de-anonymized (as spring and summer camera-ready PDFs contained author names), and all members were allowed and encouraged to review any non-conflicted candidate.

7 Reviewing for Award-Worthiness

When configuring the review form in HotCRP, the primary numeric field we used was an "award worthiness" score with a scale of 1–5 interpreted as negative, weakly negative, weakly positive, positive, and strongly positive. The associated description noted that the minimal bar we aspire to set to any ASPLOS paper is that it is well-executed and teaches us something new. Reviewers were instructed to then give bonus points if the paper met any of the following criteria: it was exciting or surprising in some sense, it is likely to stir debate, it is likely to have academic or real-world impact, it is interdisciplinary, or it is potentially of interest to the general CS community outside of ASPLOS. The description also outlined our expectations in terms of the target number of awards, as discussed earlier (Section 3).

The second numeric field we used requested reviewers to rank papers relative to the other award candidates they reviewed, with a scale of 1–3 interpreted as bottom, middle, or top third. We added this field because we were concerned that the "award worthiness" score might turn out overly negative, but this concern eventually turned out to be unfounded. For the third numeric score, we asked reviewers to assess the probability that the paper will win the ASPLOS influential paper award in 10 years, with a 1–3 scale denoting "negligible," "no idea," and "nonnegligible." There is evidence suggesting that best paper awards are poor predictors of future impact [3]. We nevertheless decided that it may be productive to try to consider this aspect when making award decisions. The fourth and final numeric field we used was the standard four-value reviewer expertise scale of no familiarity, some familiarity, knowledgeable, and expert.

In addition to the numeric scores, we asked reviewers to explain in text their award worthiness score and summarize the paper in a self-contained manner for the benefit of committee members that did not read it.

8 Review Assignment for Award Candidates

In Volume 1 and 2 of ASPLOS'24 [5, 6], we described in detail how we assigned submissions to reviewers, utilizing a constraint solver that maximizes a weighted average of topic scores, citation scores, TPMS scores [4], and predicted expertise scores. We employed the same methodology for assigning the award candidates to the members of the selection committee for review, with the following differences. (1) All committee members predicted their expertise for all non-conflicting paper candidates (instead of for a subset), because collusion was not a concern, and we wanted to achieve the best review match possible. (2) For the same reason, we extended the predicted expertise score to fully match the standard review expertise scale as specified in Table 1. (3) We did not assign candidate papers to committee members that have already reviewed them in the past in one of the three cycles, because we wanted to diversify opinion and increase the number of participants in the discussion. (4) Technically, we did not use best-matching topic information because it was added in the summer cycle and was therefore unavailable for spring papers.

Our selection committee was fairly representative of the ASPLOS community and provided good coverage in terms of research topics. But its limited size did not allow it to appropriately evaluate

Table 1. Predicted Expertise Scores used to Assign Award Candidate Papers to Selection Committee Members

predicted expertise	*meaning*
-999	I'm conflicted with this paper
-1	I'm absolutely unable to review this paper
1	I have no familiarity and my review confidence will likely be low
2	I have some familiarity
3	I may be able to provide a knowledgeable review
4	I may be able to provide an expert review

all the topics of all of the accepted papers. Thus, after we selected the award paper candidates, we identified a handful of the papers that we assessed would be problematic to confidently review in the scope of a small committee. We decided that if this turns out to be the case, then we will remove these candidates from the process. Our reasoning was not just practical, as we want the award papers to be broadly appealing for ASPLOS attendees. We deferred making the decision until after soliciting expertise predictions.

Our review budget allowed us to allocate about 6 reviews per paper candidate (because 18 candidates and 21 committee members each contributing 5 reviews give us $\frac{21\times5}{18} \approx 6$). We hoped that a majority among these 6 reviewers would be knowledgeable. We consequently decided that each candidate paper should have at least 4 committee members who expected to be able to review it knowledgeably in order for it to take part in the reviewing process. The result (which did not surprise us) was that 4 papers did not meet this bar. We filtered them out and went ahead to review the remaining 14. We budgeted 7 reviews per candidate paper, because $\frac{21\times5}{14} = 7.5$, so each committee member was assigned 4–5 candidate papers to review.

In addition, we encouraged members to add brief reviews for papers that they have already reviewed in the regular ASPLOS'24 cycles. In the description associated with the ranking score, we instructed reviewers to rank while taking their past reviews for any candidate paper into consideration.

9 Award Selection Meeting

We scheduled a 4-hour Zoom meeting for the award selection committee on April 3, 2024. Given the geographic diversity of our members, it was unfortunately impossible to select a time that was neither too early nor too late for everyone. The compromise time slot that allowed most of us to attend was early morning in Australia (5am), late evening in Europe, afternoon in the US, and too late at night for a committee member from China. We handled conflicts of interest as usual, with two of our submission chairs skillfully administering the meeting and transitioning conflicted members into and out of a breakout room.

The meeting was pleasant and easygoing, in our opinion, and we felt it was fun to get together. We scheduled 15 minutes for each paper, and our impression was that everyone had time to comfortably express their thoughts and calmly voice their opinion, including non-reviewers who occasionally participated in the discussion. At the beginning of the meeting, we outlined some tentative rules regarding how to make decisions. But the only rule that really stuck was that the discussion lead—a positive reviewer—began by introducing the paper, followed by the strongest detractor (although "strongest" was frequently too strong a term). There were some lively discussions and debates. However, the dynamic of the meeting was such that, in the end, we awarded papers only via consensus, with all of the reviewers agreeing that the candidate is worthy, possibly with one or maybe two reviewers who could be better characterized as "not objecting." Through this process, we

awarded 6 papers, which unfortunately fell at the lower end of our goal. Nevertheless, we believe that the general sense in the committee was that we made the right decisions, and that we can be proud of the outcome.

Acknowledgments

We thank Tomer Klainer and Igor Smolyar for skillfully administering the best paper selection committee meeting and for operating the conflict identification scripts. We also thank Eddie Kohler for his much appreciated and timely HotCRP support. Mostly, we thank the best paper selection committee members: Adam Morrison, Adrian Sampson, Alexandros Daglis, Andrea Lattuada, Benjamin C. Lee, Carl Waldspurger, Christopher Rossbach, Donald Porter, Emmett Witchel, Henry Hoffmann, Lieven Eeckhout, Lluís Vilanova, Marios Kogias, Martin Maas, Mingyu Gao, Rodrigo Fonseca, Sam H. Noh, Santosh Nagarakatte, Steve Blackburn, Tyler Sorensen, and Yufei Ding. Finally, we thank Emery Berger for helping us shape the best paper selection process.

References

[1] ASPLOS'24. 2024. Distinguished Reviewers. Retrieved from https://www.asplos-conference.org/asplos2024/distinguished-reviewers
[2] Emery Berger. 2017. A Guide to Selecting Distinguished Paper Awards. Retrieved from https://emeryberger.medium.com/a-guide-to-selecting-distinguished-paper-awards-7aad56ac9454
[3] Emery Berger. 2024. Best paper selection. Private email communication.
[4] Laurent Charlin and Richard Zemel. 2024. The Toronto Paper Matching System (TPMS). Retrieved from http://torontopapermatching.org
[5] Madan Musuvathi and Dan Tsafrir. 2024. ASPLOS'24 Front Matter—Volume 1: Program Chairs' Message. DOI: https://doi.org/10.1145/3617232
[6] Madan Musuvathi and Dan Tsafrir. 2024. ASPLOS'24 Front Matter—Volume 2: Program Chairs' Message. DOI: https://doi.org/10.1145/3620665
[7] Mary Lou Soffa and Mary Jane Irwin. 2009. ASPLOS XIV Welcome. DOI: https://doi.org/10.1145/1508244

Received 31 March 2026; revised 31 March 2026; accepted 1 May 2026